

Sakana AIによる「MASS」技術の全貌: マルチエージェント協調に基づく再帰的自己改善のメカニズムと影響

Gemini 3.8 Flash

発表の背景と概要: マルチエージェント自己監督(MASS)の成立

2026年10月、Sakana AIはカリフォルニア大学バークレー校(UC Berkeley)の研究陣との共同研究成果として、単一の言語モデルが自律的にマルチエージェントチームを編成・最適化し、その協調作業の履歴から自己学習を反復する革新的アーキテクチャ「MASS (Multi-Agent Self-Supervision: マルチエージェント自己監督)」を発表した¹。研究論文(arXiv:2610.12176)および実装コードはオープンに公開され、機械学習コミュニティで急速に議論を呼んでいる¹。

従来の大規模言語モデル(LLM)における事後学習や強化学習(RL)は、数学的証明やコードの構文検証のように「絶対的な正誤を決定論的に判定できる検証器(Verifier)」が存在する領域に偏重していた。しかし、創薬支援、金融市場の計量分析、オープンエンドな科学的リサーチといった複雑な実世界課題では、人間ですら単一の正解を即座に判定することは極めて困難である³。こうした検証困難なタスクにおいて、単一モデルに自己反省(Self-Correction)を行わせても、モデル自身が内在する認知バイアスや盲点を共有しているため、誤った推論を真実と見誤って正当化し、学習ループを通じてその誤謬を強化してしまう自己欺瞞のリスクが長年指摘されてきた¹。

MASSは、この「自己監督(Self-Supervision)における監督信号の欠如」という根本的課題を打破するために開発された²。単一エージェントによる長大な推論系列の生成に依存するのではなく、単一モデルの複数インスタンスに異なる責務を与えて協調チームを形成させ、エージェント間の通信と役割分担(ワークフロー)を動的に最適化する²。そして、その協調過程から得られた質の高い実行軌跡をモデル自身の重みへと蒸留(Distillation)することで、外部の正解ラベルや人間の介入なしにモデル能力を漸進的に高める再帰的自己改善(Recursive Self-Improvement: RSI)を実証した²。

アーキテクチャと自律的自己改善メカニズム

MASSのアーキテクチャ上の本質は、外部の超巨大な教師モデルに頼るのではなく、単一のオープンウェイトモデル(実験では270億パラメータの「Qwen3.6-27B」を採用)がタスク解決者、評価者、組織改善者の3役をすべて兼任する「均質型再帰的自己改善(Homogeneous RSI)」を成立させた点にある¹。

プロンプトとしてのワークフロー設計

MASSでは、マルチエージェントの組織構造が静的なプログラムとして固定されているのではなく、自然言語プロンプトとして構造化され、タスクプロンプトに動的に付加される²。これにより、モデル自身が組織構造を点検し、書き換えることが可能となっている²。

ワークフロー構成要素	定義と機能的役割	実装における具体例
役割 (Role)	各サブエージェントに割り当てられる専門的責務	データ前処理担当、予測モデル構築担当、検証監査担当
指示 (Instruction)	サブエージェントが実行すべき具体的処理内容	時系列データの欠損値補完、特定アルゴリズムによるバックテスト実行
契約 (Contract)	後続エージェントへ引き渡す成果物の仕様および検証制約	「バックテスト期間と学習期間に先読みバイアス (Data Leakage) がないことの証明データ」の添付要求
ホップ (Hop)	エージェント間の呼び出し順序、条件分岐、反復ループ	データ検証でエラーが検出された場合の前段エージェントへの差し戻し処理

この中でも特に「契約」の概念は、従来のエージェント間雑談によるコンテキストの肥大化を抑制し、厳格な入出力の整合性と自己監査を強制するアーキテクチャ上の要石となっている²。

内的ループ (Inner Loop) : ワークフローの進化探索

内的ループは、基盤モデルの重みパラメータを完全に固定した状態で進行する²。まず「実行 (Execute)」段階において、オーケストレーターとサブエージェント群が提示された候補ワークフローに厳密に従ってタスクを遂行し、コード、視覚化プロット、リサーチレポートを包含する作業環境 (ワークスペース) を生成する²。

続く「評価 (Evaluate)」段階では、同じモデルが評価者 (Evaluator) として振る舞い、現時点の最高ワークフロー (Champion) が生成した成果物と候補ワークフローの成果物をペアワイズ形式で比較判定し、採否の根拠を自然言語で言語化する²。候補ワークフローが勝利した場合、それが新たなチャンピオンとして登録される²。

最後に「改変 (Revise)」段階において、オプティマイザ (Optimizer) 役のモデルが評価者の過去のフィードバック履歴を精査し、エージェント間の役割分担や成果物の受け渡し契約、呼び出し順序の修正案を策定して次世代のワークフロープロンプトを出力する²。この一連の探索により、与えられたタスクに対して最も集合知を引き出せる組織構造が自律的に洗練されていく²。

外的ループ (Outer Loop) : 集合知のパラメータ蒸留

内的ループによって特定されたチャンピオン・ワークフローの成果は、外的ループによってモデルそのものの基礎知能へと還元される²。

システムはまず、最適化されたワークフローを乱数シードを変えて18回試行し、生成されたワークスペース群をモデル自身による相互評価とBradley-Terryモデルによって順位付けし、最上位16件の

優れた実行軌跡を抽出する²。ここで極めて重要な処理として、抽出された対話ログからオーケストレーターの初期プロンプトに含まれていた「明示的なワークフロー指示テキスト」が完全に除去される²。この前処理を施すことで、モデルは外部からチーム指示を与えられなくとも、課題文そのものから直接的に「自律的な課題分解」「他者への委任」「厳密な検収」「成果物の統合」を行う振る舞いを内在化できるようになる²。

最終的に、オーケストレーターの統括対話とサブエージェントの局所的実行ログの双方が教師データとして組み込まれ、LoRA (Low-Rank Adaptation) を介して共有モデルのパラメータが更新される¹。パラメータが更新されたモデルは、次世代の学習サイクルにおいて再びタスク実行者、評価者、オプティマイザの全役職に就任し、さらに一段階高い次元での自己改善ループを開始する²。

実証実験とベンチマーク結果の分析

実験では、ベースモデルとしてQwen3.6-27Bが用いられ、金融(株式戦略バックテスト)、ロボティクス、薬理学を含む12件の高度な合成研究タスク環境で検証が行われた¹。改善プロセスを外部から客観的に測定するため、学習ループからは完全に遮断された外部フロンティアモデル(GPT-5.5およびClaude Opus 4.8)がブラインド評価者として配備された¹。

ベンチマーク評価と性能推移

ベースモデル($\mathcal{L}^{(0)}$)、第1世代改善モデル($\mathcal{L}^{(1)}$)、第2世代改善モデル($\mathcal{L}^{(2)}$)の比較において、MASSIはオープンエンドな研究タスク群で大幅な性能向上を示した²。

評価指標・ベンチマーク	ベースモデル対比の改善度	詳細および備考
未出の研究タスク(Held-Out Suite)	勝率 69.9% (第2世代)	第1世代は53.9%、第2世代は第1世代に対しても60.8%の勝率を記録 ¹
MLR-Bench(機械学習リサーチ)	1.60倍 (出力トークンあたり)	12タスクの平均総合スコアが 1.58 ± から 2.40 ± へ向上 ²
DSBench(データサイエンス)	1.35倍 (出力トークンあたり)	複雑なデータパイプライン処理の効率化を実証 ²
AstaBench (E2E-Hard)	1.28倍 (出力トークンあたり)	エンドツーエンドの複合推論におけるスケーリング向上 ²
ScienceAgentBench(科学研究)	1.20倍 (出力トークンあたり)	専門的仮説検証プロセスの洗練を確認 ²
Terminal-Bench 2.0(環境)	0.96倍 (軽微な低下)	シェル環境特有の厳格な構

操作)		文実行における転移の非対称性 ¹
SWE-bench Verified (ソフトウェア工学)	0.94倍 (軽微な低下)	決定論的なコード修正タスクへの汎化課題を示唆 ¹

オープンエンドなリサーチ系ベンチマーク群において出力トークンあたりのスコアが1.2~1.6倍に向上した一方で、決定論的な単一解が求められるSWE-bench VerifiedやTerminal-Bench 2.0ではスコアが基準値を下回る結果となった¹。研究チームはこの非対称性について、訓練タスク群が学術的・探索的なリサーチ領域に特化していたため生じた分布の偏りであり、訓練データセットにソフトウェア工学タスクを補完することで解消可能であると論じている¹。

データ効率性: マルチエージェント軌跡の優位性

本研究における極めて重要な理論的発見は、自己学習の教師データとして「マルチエージェントの協調軌跡」が「単一エージェントの長大軌跡」よりも本質的に優れた学習効率を誇るという点の実証である¹。

学習手法とデータ規模	消費された教師トークン数	ベースモデルに対する勝率
単一エージェント軌跡による学習 (S^{++})	4,200万トークン (42.00M)	64.0% ¹
マルチエージェント軌跡による学習 (M^{+})	2,882万トークン (28.82M)	68.3% [cite: 1, 5]

単一エージェントが単一の長大な文脈の中で計画立案、ツール呼び出し、自己反省を混在させて生成したログは、文脈の肥大化に伴うアテンションの希釈や推論ノイズを不可避免的に含む⁴。これに対して、オーケストレーターが抽象度の高い統合を担い、サブエージェントが契約に基づき限定的な局所タスクを完結させるという階層的模倣学習の枠組みは、情報の冗長性を劇的に削減する⁴。結果として、約31%少ないトークン数でありながら4.3ポイント高い勝率を叩き出し、マルチエージェント協調のログが学習信号として極めて優良であることを証明した¹。

技術的新規性と理論的洞察: 創発するメタ能力

MASSの実験結果は、モデルの個別タスク解決能力の向上にとどまらず、再帰的自己改善のメカニズムに関する従来の理論的仮説を覆す洞察をもたらした。

役割汎化 (Role Generalization) の成立

モデルは「マルチエージェントによるタスク解決軌跡」のみを用いて事後学習されたにもかかわらず、パラメータ更新後にはタスク実行能力のみならず、ワークフローを設計するオプティマイザとしての能力、および成果物を評価する評価者としての能力の双方が自律的に向上した²。

第1世代のオプティマイザが立案したワークフローをベースモデルに実行させた場合、ベースモデル自身が設計した初期ワークフローに対して64%の勝率を記録した²。さらに、モデル内部の自己評価

と、外部の強力なフロンティアモデル（GPT-5.5、Claude Opus 4.8）との判定一致率は、ベースモデルの73%から改善サイクルを経て93%へと到達した¹。これは、協調的なタスク解決の過程を学習すること自体が、高次のメタ認知的判断力や組織プロセスの設計能力を汎化的に刺激することを示している²。

RSIIにおける真のボトルネックの特定

従来のRSI研究においては、モデル自身が評価を下す際に「自己評価能力の不足（Evaluator Bottleneck）」がシステムの崩壊を招く最大の要因であると広く信じられてきた²。しかし、MASSにおけるアブレーション実験は、改善サイクルの真の障壁が「評価者」ではなく「オプティマイザ（Optimizer）」に存在することを突き止めた²。

評価者のみを外部の最先端モデルであるGPT-5.5に差し替えても、改善ループの進行速度やワークフローの品質はほとんど向上しなかった²。高度な評価者が極めて的確で詳細な批判的フィードバックを提供しても、低能力のオプティマイザはその指摘を咀嚼し、具体的な組織構造や契約の改善案へと落とし込むことができなかったためである²。限られた計算予算の制約下で再帰的自己改善を推進する場合、評価器の肥大化よりも、プロセスを再設計するオプティマイザの能力強化に計算資源を傾斜配分すべきであることが実証された²。

テスト時計算量の重みパラメータへの内面化

近年のLLM応用においては、推論時（テスト時）に多数のエージェントを稼働させて合議や検証を行わせる「テスト時スケーリング」が主流となってきたが、これには膨大な推論コストとレイテンシの増大が不可避であった。MASSの成果は、推論時に動員されるマルチエージェント協調という計算集約的な振る舞いを、学習を通じて「モデルの重みパラメータ」へと直接焼き込めることを示した²。事後学習を経たモデルは、推論時に明示的なエージェントフレームワークが与えられずとも、課題の性質に応じて自律的かつ暗黙的に自己委任や検証プロセスを展開する能力を獲得する²。

学術界・産業界の反響と技術的論争

MASSの発表は、Hacker NewsやRedditなどの開発者コミュニティ、ならびにAI研究者の間で瞬く間に大きな議論を巻き起こした⁷。

学術コミュニティでは、単なるプロンプト制御や外部オーケストレーションツールによるエージェント運用を超え、エージェントシステムの実行成果をLoRAを通じてパラメータに回帰させる完全な閉ループを構築した点に高い賞賛が集まった⁸。さらに、莫大なAPI利用料を消費するフロンティアモデルではなく、27Bクラスのオープンモデルを用いて単一GPU環境でも再現可能な規模で自己改善ループを成立させた点は、オープンソースAIエコシステムや機密保護を最優先するエンタープライズにとって、現実的かつ持続可能なスケーリングの道筋を示したと高く評価されている¹¹。

一方で、自己評価モデルの構造的宿命に関する根強い懐疑論も噴出している。モデル自身が実行と評価を兼任するシステムは、本質的にグッドハートの法則（指標が目標になると適切な指標でなくなる現象）や報酬ハッキングの罠に陥りやすく、人間にとって真に有用な推論ではなく、モデル独自の閉じた評価基準に過剰適合しているだけではないかという懸念である⁸。また、決定論的な検証が不可欠なソフトウェア工学タスク（SWE-bench Verifiedで0.94倍）でのスコア低下は、探索的なリサーチ能力の獲得と引き換えに厳格な構文処理能力が損なわれる「破滅的忘却」のトレードオフを想起させると警戒されている¹。さらに一部の研究者からは、今回の成果は事前学習データに潜在していた知識の再構成にすぎず、人類の未踏領域を拓く真の超知能への離陸と呼ぶには時期尚早であるとの冷静な見方も提示されている¹¹。

安全性、アライメント、今後の展望

自律的に組織を改編し自己学習を反復するマルチエージェント技術の台頭は、AIアライメントと安全保障の観点からも極めて重大な論点を投げかけている。

均質型RSIIにおける最大の破滅的シナリオは「共通の認知バイアス(Shared Blindspots)」の自己増幅である¹。モデルが特定の論理的破綻や誤謬を「優れた成果物」と誤認した場合、その誤謬を正当化するワークフローがチャンピオンとして選定され、誤った前提に立った軌跡がパラメータに学習されてしまうため、世代を経るごとにモデルが独自の妄想を強固に確信していく危険性を孕んでいる¹。Sakana AI自身も、AIが人間の専門知識を超える領域へ自律進化する過程においては、こうした潜在的暴走を外部から確実に検知し、強制介入できるセーフガード機構の維持が不可欠であると強調している³。

この安全上の課題に対し、本研究プロトコルが提示した「外部評価モデル(GPT-5.5、Claude Opus 4.8)を学習ループから完全に切り離れたサンドボックスに隔離し、客観的監査役として常時並行稼働させる手法」は、今後の自律的AI開発における標準的な安全監査アーキテクチャのひな型となる可能性がある¹。

産業界への波及効果としては、企業におけるAIエージェント導入のパラダイムが根本から覆される可能性が高い。従来のマルチエージェント構築では、システムエンジニアが各エージェントの役割分担やメッセージのやり取りを手作業で泥臭く設計・調整する必要があった。しかしMASSの手法が確立されれば、人間は最終的な達成目標と成果物の契約制約(Contract)を定義するだけで、AI自身が最適な組織編成を試行錯誤し、業務マニュアル(ワークフロー)を自律的に洗練させ、最終的にはその業務ノウハウを専用の軽量モデルに吸収させることが可能となる²。金融分析、戦略リサーチ、特許調査、先端材料探索といった高度な専門業務において、エージェントパイプラインの構築・保守コストが劇的に圧縮される未来が現実味を帯びている⁵。

結論

Sakana AIとUC Berkeleyが提示した「MASS」は、単一の27Bモデルを用いて、組織編成、実行、自己評価、そしてLoRAを通じたパラメータへの蒸留という再帰的自己改善のサイクルを完全に具現化した画期的なマイルストーンである¹。単一エージェントの長大な推論よりも、役割と契約によって分割されたマルチエージェントの協調軌跡の方がはるかに高い学習効率をもたらすという発見は、今後の基盤モデルの事後学習パラダイムに決定的な転換を迫るものである¹。

決定論的タスクへの汎化の非対称性や、閉ループ学習が孕む自己欺瞞の抑制といった技術的・倫理的課題は依然として立ちはだかっているものの、MASSが示した「集合知の自己監督による自己改善」というアプローチは、AIが人間の監督能力の限界を超えて知能を拡張していくための強固な足がかりを築いたといえる¹。

引用文献

1. Sakana AIがMASSを発表、Qwen3.6-27Bが自身のマルチ ... - jls42.org,
<https://jls42.org/ja/news/ia-actualites-11-oct-2026>
2. Multi-Agent Self-Supervision (MASS) - Sakana AI, <https://pub.sakana.ai/mass/>
3. Recursive Self-Improvement through Multi-Agent Self-Supervision,
<https://sakana.ai/mass/>
4. Recursive Self-Improvement through Multi-Agent Self-Supervision,

- <https://arxiv.org/html/2610.12176v1>
5. Recursive Self-Improvement through Multi-Agent Self-Supervision,
<https://www.alphaxiv.org/abs/2610.12176>
 6. Paper Skygest - Bluesky,
<https://bsky.app/profile/paper-feed.bsky.social/feed/preprintdigest>
 7. AIデイリー | Anthropicがエージェントの制御問題で联网停止,
<https://www.communeify.com/ja/blog/ai-daily-2026-10-11/>
 8. Recursive Self-Improvement Is Here. Sakana AI Demonstrates a,
https://www.reddit.com/r/singularity/comments/1x3khul/recursive_selfimprovement_is_here_sakana_ai/
 9. Sakana AI's MASS: a model organizes copies of itself... - daily.dev,
<https://daily.dev/posts/sakana-ai-s-mass-a-model-organizes-copies-of-itself-into-teams-then-trains-on-what-worked-gqk05njis>
 10. alphaXiv: Explore, <https://www.alphaxiv.org/>
 11. <https://news.ycombinator.com/item?id=50048522>
 12. <https://hermes-ai.net/news/sakana-ai-s-mass-trains-one-model-to-run-its-own-agent-team/>
 13. Sakana AI Blog, <https://sakana.ai/blog/>
 14. 日本から、フロンティアAIを - Sakana AI, <https://sakana.ai/ja/>